UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

March 8, 2001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES R-40

MEMORANDUM FOR Maureen Lynch
Assistant Division Chief, Coverage Measurement Processing
Decennial Statistical Studies Division

From: Donna Kostanich DK
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Thomas Mule TM
Sample Design Team

Subject: Accuracy and Coverage Evaluation: P-Sample and E-Sample Weights

I. INTRODUCTION

This memorandum provides specifications for applying the Accuracy and Coverage Evaluation (A.C.E.) survey sampling weights to the Missing Data Input Files. These specifications cover the assignment of the P-sample and the E-sample trimmed weights for each A.C.E. sample cluster to the Missing Data Input files and the Sample Design File.

The unbiased weights reflect several A.C.E. sampling operations: the listing sample selection, A.C.E. block cluster reduction, small block subsampling, large block subsampling for the P-sample. These weights were calculated during the large block cluster subsampling process. The E-sample weights are the results of all of the above operations plus the E-sample identification, where they were calculated.

The Sample Design team might have to trim the P-Sample and E-Sample weights because the weighted influence that a cluster contributes to an estimate may be extremely high. By examining After Person Follow Up matching results, the Sample Design team will identify and, if necessary, trim the weights of highly influential clusters. Identifying and trimming the weights in a prespecified manner will be conducted by the Sample Design staff. See Q-26 for an overview of the weight trimming methodology.

The Sample Design team will create version 7 of the Sample Design File. The Processing Support team will use this file to assign the weights to the Missing Data files.

The specification is ordered as follows:

Section II gives the assumptions and definitions.
Section III lists the input and output files for this process.
Section IV specifies how to apply the weights to the Missing Data Input files.

The Sample Design team is responsible for verifying the calculation of the weights and ensuring they are properly assigned to the Missing Data Input Files. The scope of this verification of the Missing Data Input Files will be limited to the weights.

Note that this work is complete. This specification reflects the process as it was actually implemented and includes any required changes to the working draft.

II. ASSUMPTIONS AND DEFINITIONS

A. The P-sample weights will be assigned to the Missing Data P-sample Housing Unit Input file while the E-sample weights will be assigned to the Missing Data E-sample Person Input file. No weights are assigned at this time to the Missing Data P-sample Person Input file. Part of the Missing Data system is to generate non-interview factors and targeted extended search adjustments to generate final weights for the Dual System Estimate.

B. These weights can not be assigned until the weight trimming operation is completed.

C. The E-sample probability of selection code (ESPS): The ESPS code was assigned to each E-sample housing unit during E-sample Identification. Each cluster can have up to two E-sample weights. The ESPS code identifies which weight to assign to each person. See R-39 for more information.

III. FILES

A. INPUT FILES

1. Sample Design File (SDF): Version 7 of the A.C.E. Design file reflects all A.C.E. sampling operations through weight trimming. This file is at the block cluster level and contains 29,717 records, which is one record for each block cluster selected for the first step of the listing sample selection.

Refer to Attachment A for the complete file layout. This specification will utilize the following variables from version 7 of the file during processing.

Variable Description	Name	Position
A.C.E. block cluster number (including check digit)	CLUSTER	21-26
Trimmed weight for P-sample HUs	TRIMWTP	597-608
Trimmed weight for E-sample HUs with an ESPS code of 1	TRIMWTE1	610-621
Trimmed weight for E-sample HUs with an ESPS code of 2	TRIMWTE2	623-634

2. Missing Data P-Sample Housing Unit Input File: The P-Sample trimmed weights will be assigned to housing units on this file. The P-Sample trimmed cluster weight is assigned to each housing unit in the cluster. Refer to Q-40 for the layout.

3. Missing Data E-Sample Person Input File: The E-Sample trimmed weights will be assigned to people on this file. There are indicators on this file to link the correct E-Sample trimmed weight to each person. Refer to Q-40 for the layout.

B. OUTPUT FILES

The Sample Design Team needs access to the following Missing Data Input files to verify the weights have been correctly assigned.

- E-Sample Person
- P-Sample Housing Unit
- P-Sample Person

IV. ASSIGNING WEIGHTS TO THE MISSING DATA INPUT FILES

A. Read Input Files

1. For each cluster remaining in sample, obtain the trimmed P-sample and E-sample cluster weights from the Sample Design File.

Variable Description	Name
A.C.E. block cluster number (including check digit)	CLUSTER
Trimmed P-sample Cluster Weight	TRIMWTP
Trimmed E-sample Cluster Weight for ESPS code of 1	TRIMWTE1
Trimmed E-sample Cluster Weight for ESPS code of 2	TRIMWTE2

2. For each housing unit, obtain the cluster number from the P-sample Housing Unit Missing Data Input File.

Variable Description	Name
A.C.E. block cluster number (including check digit)	CLUSTER

3. For each person on the E-Sample Missing Data Input file, obtain the cluster number and the E-sample probability of selection code.

Variable Description	Name
A.C.E. block cluster number (including check digit)	CLUST
E-sample Probability of selection code	ESPS

B. Link the Weights to the Missing Data Input Files

The trimmed cluster weights need to be linked to the appropriate P-sample housing units and E-sample persons on the Missing Data Input Files. Link the weights as follows:

1. E-sample Person File

   Assign each record a trimmed weight based on block cluster code and ESPS code. Assign each record in a block cluster with an ESPS code of 1 an E-sample weight equal to TRIMWTE1. Assign each record in a block cluster with an ESPS code of 2 an E-sample weight equal to TRIMWTE2.

2. P-sample Housing Unit File

   Assign the P-sample trimmed cluster weight (TRIMWTP) to each housing unit in the cluster based on block cluster value.

3. P-sample Person file

   No weights are assigned to the P-sample Person Input file. The Missing Data Process assigns the P-sample person weights.

References

R-39 Memorandum for Lynch from Kostanich, "Accuracy and Coverage Evaluation Survey: Specification for E-Sample Identification," December 7, 2000.

Q-26 Memorandum for Hogan from Kostanich, "Accuracy and Coverage Evaluation: Weight Trimming Plan," July 20, 2000.

Q-40 Memorandum for Lynch through Kostanich from Cantwell "Census 2000: Specifications for Data Requirements for A.C.E. Missing Data Input and Output Files," January 19, 2001.

cc: DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List
ACE Implementation Team/Statistical Design Team Leaders List
DSSD Sample Design Team

Sample Design File

The Sample Design File contains one record per block cluster selected during the listing sample selection. If the block clusters falls out of sample during the second step of sampling or during small block subsampling, the remaining variables are blank. The layout for the Sample Design File is as follows:

Variable Description	Name	Places	Source
Census Region	REGION	1	UN
Census Division	DIV	2	UN
State code	STATE	3-4	UN
County code	COUNTY	5-7	UN
Local census office	LCO	8-11	CS
Interim Tract (Pseudo Tract)	ITRACT	12-17	BC
Current Sample Indicator	CSI	19	UO
A.C.E. block cluster number	CLUST	21-25	CS
Check Digit	DIGIT	26	CS
Geography block cluster number	GCLUST	28-32	BC
List/Enumerate Indicator	LEIND	33	BC
Type of Enumeration Area Recode	TEACR	34	CS
Type of Enumeration Area Group Revised	TEAGR	35	SB
Type of Enumeration Area group	TEAG	36	BC
Number of HUs used for sample design	NHU	37-41	BC
Number of MAF HUs	NHUM	43-47	BC
Number of 1990 HUs	NHU90	49-53	BC
Sampling Stratum 1 = Small 2 = Medium 3 = Large 4 = American Indian Reservation	SS	55	UN
American Indian Country Indicator 0 = No American Indian Country 1 = American Indian Reservation/trust land 2 = Tribal Jurisdiction Area/ Alaska Native Village Statistical Area/ Tribal Designated Statistical Area	AICIND	56	BC
Demographic/Tenure Group code	DTCODE	57-58	UN
Demographic/Tenure Group label	DTLABEL	59-60	UN
Estimated Urbanicity of block cluster 1 = Urban Area with population ≥250,000 2 = Other Urban Area 3 = Non-Urban Area	ECLUSURB	62	UN
Size Category 1=Small (0-2 hus) 2=Medium (3-79 hus) 3=Large (80+ hus)	SIZCAT	63	UN
Additional space		64-91	

Variable Description	Name	Places	Source
First step index number	INDEX1	92-99	CS
Listing sample selection indicator 1 = Selected	BC1	101	CS
Random Start for listing sample selection	RS1	103-113	UN
Take-every for listing sample selection	TE1	115-125	UN
Second step listing sample selection indicator 0 = Not Selected 1 = Selected	BC2	127	CS
Random Start for second block cluster sampling	RS2	129-139	CS
Take-every for second block cluster sampling	TE2	141-151	CS
Unbiased weight after block cluster sampling	WEIGHTBC	153-164	CS
Additional space		165-175	
Preliminary Number of HUs on the Independent List	NHUILP	176-180	AR
Number of Housing Units On the DMAF	NHUDMAF	182-186	AR
Demographic Code 1 = Minority 2 = Non-Minority 3 = Puerto-Rico	DEMCODE	188	AR
Consistency Code 1 = Low Inconsistent (IL significantly smaller than DMAF) 2 = Consistent 3 = High Inconsistent ((IL significantly larger than DMAF)	CONCODE	189	AR
A.C.E. Reduction Stratum	ARST	190-191	AR
A.C.E. Reduction Indicator 0 = Not Selected 1 = Selected	ACERED	193	AR
Random Start for A.C.E. Reduction	RSAR	195-205	AR
Take-every for A.C.E. Reduction	TEAR	207-217	AR
Unbiased weight after A.C.E. reduction	WEIGHTAR	219-230	AR
Collapsing Flag	COLFLAG	232	AR
A.C.E. Reduction Index Number	INDEXR	234-241	AR
Number of Housing Units On the December 1999 DMAF (Initial)	NHUDMAFI	243-247	AR
Additional space		250-300	
Number of HUs on the Independent List Revised	NHUILR	295-299	SB
Number of HUs on the Independent List	NHUIL	301-305	SB
Small Block Cluster Subsampling Stratum	SBCSS	306-307	SB
Small Block Subsampling Indicator 0 = Not Selected 1 = Selected	SB	308	SB
Random Start for Small Block subsampling	RSSB	310-320	SB
Initial take-every for Small Block subsampling	ITESB	322-332	SB
Unbiased weight for A.C.E. cluster	WEIGHTC	334-345	SB
Larger of the DMAF and IL HU count	LARGERHU	347-351	SB
Final take-every for Small Block subsampling	FTESB	352-362	SB
Additional space		363-370	
Relisted Block Cluster Flag 0 = Not Relisted, 1 = Relisted	RELIST	371	LB
Number of total hus in block cluster	NHUEL	373-377	LB
Number of A.C.E. hus in cluster	NHUELA	379-383	LB

Variable Description	Name	Places	Source
Number of supplemental hus in cluster	NHUELN	385-389	LB
Large Block Cluster EL subsampling code 1 = NHUELI< 80 hus, 2 = NHUELI ≥ 80 hus	ELLBSUB	391	LB
Random Start for Large Block subsampling	RSLB	393-403	LB
Take-every for Large Block subsampling	TELB	405-415	LB
Number of segments in block cluster	NSEG	417-418	LB
Number of segments selected in block cluster	NSEGSAM	420-421	LB
Day of Arrival	DAY	423-424	LB
Final Cluster Order Number	CON	431-434	LB
Number of total hus for interview in block cluster	NINT	436-440	LB
Unbiased weight for P-sample HUs	WEIGHTP	442-453	LB
Number of Assignments in block cluster	NA	455-456	LB
Final Sampling Strata	FINSS	458-464	LB
Additional space		465-490	
Number of HCUF HUs in block cluster with an ESPS code of 1	NHUCUF1	491-495	ES
Number of HCUF HUs in block cluster with an ESPS code of 2	NHUCUF2	497-501	ES
Number of HCUF HUs in block cluster	NHUCUF	503-507	ES
Number of HCUF HUs in selected segments with an ESPS code of 1	NHUCUFS1	509-513	ES
Number of HCUF HUs in selected segments with an ESPS code of 2	NHUCUFS2	515-519	ES
Number of HCUF HUs in selected segments of a block cluster	NHUCUFS	521-525	ES
E-Sample Identification cluster category 1 = NHUCUF < 80 2 = NHUCUF ≥ 80 and NHUCUFS < 80 3 = NHUCUF ≥ 80 and NHUCUFS ≥ 80 4 = NHUCUF ≥ 80 and RELIST = 1 5 = NHUCUF ≥ 80 and List/Enumerate 6 = NHUCUF ≥ 80 and is a special cluster	EICC	527	ES
Random Start for E-sample subsampling	RSES	529-539	ES
Take-every for E-sample subsampling	TEES	541-551	ES
Number of E-sample HUs in block cluster with an ESPS code of 1	NHUES1	553-557	ES
Number of E-sample HUs in block cluster with an ESPS code of 2	NHUES2	559-563	ES
Number of E-sample HUs in block cluster	NHUES	565-569	ES
Unbiased weight for E-sample HUs with an ESPS code of 1	WEIGHTE1	571-582	ES
Unbiased weight for E-sample HUs with an ESPS code of 2	WEIGHTE2	584-595	ES
Trimmed weight for P-sample HUs	TRIMWTP	597-608	WT
Trimmed weight for E-sample HUs with an ESPS code of 1	TRIMWTE1	610-621	WT
Trimmed weight for E-sample HUs with an ESPS code of 2	TRIMWTE2	623-634	WT
Additional Space		636-675	

Variable Description	Name	Places	Source
Number of confirmed A.C.E. housing units not found in the census	CURCI	676-680	TES
Number of unconfirmed A.C.E. housing units not found in the census	CURUI	682-686	TES
Number of census housing units geocoded to the wrong census block	CURGE	688-692	TES
Targeted extended search selection type	TESSELECT	694	TES
Targeted extended search selection flag	TESFLAG	696	TES
Random Start for the targeted extended search	RSTES	698–709	TES
Take-every for the targeted extended search	TETES	710–721	TES
Targeted Extended Search Index Number	TESN	722-727	TES
Additional Space		728-750	

Source Codes

AR: A.C.E. Reduction
BC: Block Clustering
CS: Block Cluster Sampling
ES: E-sample Identification
LB: Large Block Subsampling
SB: Small Block Subsampling
UN: Universe File Creation
UO: Updated for each operation
WT: Weight Assignment
TES: Targeted Extended Search